




Lossless Compression Method for the Magnetic and Helioseismic Imager (MHI) Payload

Li-Yue Tong^{1,2,3}, Jia-Ben Lin^{1,2}, Yuan-Yong Deng^{1,2,3}, Kai-Fan Ji^{3,4} , Jun-Feng Hou^{1,2,3}, Quan Wang^{1,2}, and Xiao Yang^{1,2}

¹ National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China; jiabenlin@bao.ac.cn

² Key Laboratory of Solar Activity and Space Weather, National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China

³ School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China

⁴ Yunnan Observatories, Chinese Academy of Sciences, Kunming 650216, China

Received 2024 January 12; revised 2024 February 3; accepted 2024 February 19; published 2024 April 10

Abstract

The Solar Polar-orbit Observatory (SPO), proposed by Chinese scientists, is designed to observe the solar polar regions in an unprecedented way with a spacecraft traveling in a large solar inclination angle and a small ellipticity. However, one of the most significant challenges lies in ultra-long-distance data transmission, particularly for the Magnetic and Helioseismic Imager (MHI), which is the most important payload and generates the largest volume of data in SPO. In this paper, we propose a tailored lossless data compression method based on the measurement mode and characteristics of MHI data. The background out of the solar disk is removed to decrease the pixel number of an image under compression. Multiple predictive coding methods are combined to eliminate the redundancy utilizing the correlation (space, spectrum, and polarization) in data set, improving the compression ratio. Experimental results demonstrate that our method achieves an average compression ratio of 3.67. The compression time is also less than the general observation period. The method exhibits strong feasibility and can be easily adapted to MHI.

Key words: methods: data analysis – techniques: image processing – Sun: magnetic fields – Sun: photosphere

1. Introduction

The solar polar regions are vital in controlling solar activity and driving space weather, but remain as the least-known mysterious territory on the Sun. So far, some polar exploration activities have been carried out, such as Ulysses (Wenzel et al. 1992) and Solar Orbiter (Müller et al. 2020), but there have been no frontal imaging observations on the solar polar regions (Nandy et al. 2023). The Solar Polar-orbit Observatory (SPO) proposed by Chinese scientists will directly image the solar polar regions in an unprecedented way with a spacecraft traveling in a large solar inclination angle ($\geq 80^\circ$) and a small ellipticity. By obtaining the magnetic field and flow field with high precision, and combining multi-band remote-sensing and in situ measurements, SPO will make breakthrough on the following top-level scientific objectives: to unveil the origin of the solar magnetic activity cycle that shapes the living environment of human beings, to unveil the origin of the high-speed solar wind that connects the Sun and celestial bodies in the solar system, and to construct data-driven global heliospheric numerical models which serves as the foundation for space weather prediction (Deng et al. 2023).

SPO will be equipped with nine payloads, of which the Magnetic and Helioseismic Imager (MHI) is the most critical payload. MHI will provide high-resolution and high-sensitivity measurements of the photospheric vector magnetic field and

Doppler velocity with cadence of 15 and 1 minutes, respectively. However, due to the particular orbit of SPO, efficient scientific data compression and transmission schemes are highly concerned. The first period of north polar observation for SPO ($\geq 55^\circ$) will generate about 40 Tb (about 23.1 Tb from MHI) of scientific data, while the available data transmission capacity may be about 12 Tb. Limited by bandwidth, the data have to be compressed to less than 3.5 (Deng et al. 2023). Therefore, exploring high-performance data compression methods is of great importance for MHI.

Compression algorithms for general purpose, such as Huffman coding, run-length encoding, and LZW encoding, are widely applied and can achieve high compression ratios. However, other factors including the absence of patent issues, low complexity, and simple hardware implementation also need to be considered for space payloads. Consequently, compression algorithms for space payloads typically adopt some international space compression standards as a blueprint to be further refined and improved upon.

There are two distinct stages for the high-performance data compression in solar space exploration. In the early stage, limited by the technology, lossy compression with higher compression ratio was chosen after weighing limited bandwidth, massive data transmission, and tolerable data distortion. For example, the Transition Region and Coronal Explorer (TRACE), Solar and

Heliospheric Observatory (SOHO), Solar Dynamics Observatory (SDO), and Hinode used high-quality lossy JPEG compression to reduce data transmission bandwidth dependence (Fischer et al. 2016). In recent years, the lossless compression has been taken into consideration due to the rapid development of transmission technology and the relentless pursuit of high-resolution solar observations. Currently, most payloads refer to the Lossless JPEG image compression standard coding or the data compression standard recommended by the Consultative Committee for Space Data Systems (CCSDS) data compression working group. The Lossless JPEG includes two typical subtypes: JPEG2000 (ITU-T 2002) and JPEG-LS (ITU-T 1998), the Full-disk MagnetoGraph (FMG) onboard the Advanced Space-based Solar Observatory (ASO-S) applies the JPEG2000 in the data compression (Deng et al. 2019). The CCSDS recommended standard adopts a two-dimensional discrete wavelet transform and slight plane coding for images (CCSDS 2017), and uses Golomb-Rice encoding (Golomb 1966) for arbitrary data streams (CCSDS 2020). Liu (2018) compared the general compression algorithm and the Golomb-Rice algorithm to compress the full-disk solar photosphere image, and their maximum compression ratios were 2.053 and 2.084, respectively.

However, the above methods do not consider eliminating the inherent redundancy in prior knowledge, which is potential to significantly enhance data lossless compression efficiency. In this article, we address this issue by removing background information unrelated to the solar disk from the raw data, as well as invariant features among consecutive frames and other forms of redundant prior knowledge. Since the background information is irrelevant to our research objectives and given that other forms of prior knowledge redundancy are reversible and can be reconstructed during decompression, we anticipate that this substantial reduction of redundancy will result in a marked improvement in lossless data compression performance, particularly suitable for MHI applications.

Section 2 elaborates on the measurement mode of the solar magnetic field and Doppler velocity field for MHI, and proposes a compression strategy based on the measurement mode; Section 3 details the compression method; in order to verify the superiority of the method, Section 4 compares the compression ratios of different compression methods, evaluates the compression time of our method through experiments, and discusses the factors affecting the effectiveness and the feasibility of spatialization; finally, Section 5 gives a summary.

2. Characteristic Analysis of MHI Data

2.1. The Measurement Mode of the Solar Magnetic Field and Doppler Velocity Field for MHI

Presently the solar magnetic field is mainly measured based on the Zeeman effect. MHI will carry out measurements at several wavelength positions around the Zeeman-sensitive photospheric spectral line and its nearby continuum using

narrowband filters, and analyze the Stokes polarization parameters (I , Q , U , and V) of the incident light at each spectral position using the polarization modulator. With these observations, we can derive the vector magnetic field. Therefore, the magnetic field measurement by a solar telescope is essentially a polarization measurement, that is, the measurement of Stokes parameters I , Q , U and V .

The polarization measurement of MHI uses a differential mode. Taking the Q parameter as an example, the detected intensities $P(I+Q)$ and $P(I-Q)$ are obtained through polarization modulation, the difference between $P(I+Q)$ and $P(I-Q)$ is used to obtain the Q parameter, and the sum of them is I . In order to improve the signal-to-noise ratio, short exposure, alternating sampling, and integration for $P(I+Q)$ and $P(I-Q)$ are applied (Ai & Hu 1986). The equation for obtaining Stokes Q , U , and V is :

$$S = \frac{\sum P(I+S) - \sum P(I-S)}{2}, \quad (S = Q, U, V) \quad (1)$$

After measuring the imaging intensity at multiple wavelength positions near the center of the working spectral line, the offset from the line center is calculated by fitting the spectral contour at multiple wavelength points. The solar line-of-sight Doppler velocity can be obtained according to the proportional relationship between Doppler velocity and the drift of the line center (Zhang et al. 2007).

The compression of original data for MHI is essentially compressing solar polarization data sets $P(u, v, \lambda_c + \Delta\lambda, I \pm S)$. u and v denote the indices for the width and height of the image, λ_c , $\Delta\lambda_{n-1}$, and $I \pm S$ represent the center wavelength, the offset and polarization modes, respectively (Figure 1).

2.2. Analysis of Compression Strategies from Prior Knowledge

From our particular observation data, we extract five points of characteristics, which can be utilized to significantly improve the data compression efficiency.

1. For full-disk solar images, the effective information is concentrated within the solar disk, and appropriate templates can be used to extract the information, thereby reducing overall data volume (Figure 2(a)).
2. Given the large-scale hemispherical shape of full-disk intensity, a specific predictive coding can be devised to eliminate spatial redundancy and to lower the pixel intensity values of single-frame data (Figure 2(b)).
3. The weak intensity difference between $P(I+S)$ and $P(I-S)$ can be retained to achieve a superior compression effect (Figure 2(c)).
4. For the same wavelength of QUV data, the sum of $P(I+S)$ and $P(I-S)$ is approximately invariant: $P(I+Q) + P(I-Q) \approx P(I+U) + P(I-U) \approx P(I+V) + P(I-V)$.

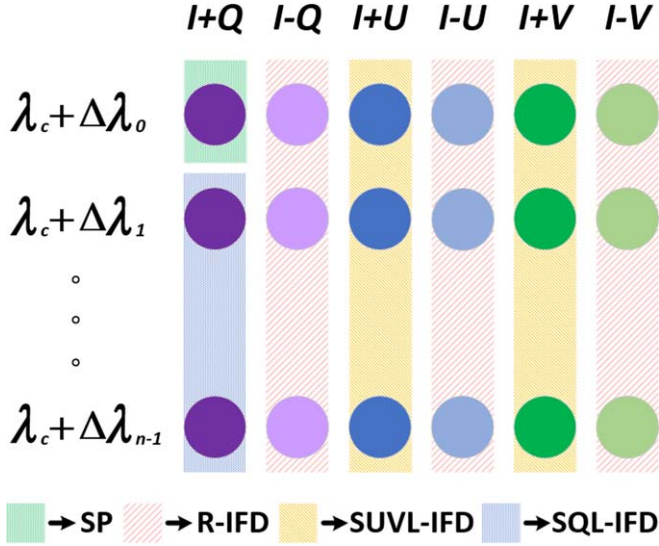


Figure 1. Overview of spectral scanning data predictive coding.

5. In Doppler velocity field observation, there exhibit large-scale intensity variations across different wavelengths (Figure 2(d)).

Referred to compression theory, the above five characteristics can be summarized as background removal, elimination of spatial redundancy in a single frame, and elimination of both polarized and spectral redundancy in continuous frames.

3. Design of Lossless Compression Method

Figure 3 shows the whole process of compression and decompression. First, we design an extractor that can effectively remove most of the external background from the solar image, and then perform dimensionality reduction on the extracted solar information to facilitate subsequent processing. Next, a secondary pre-pixel (SP) predictive coding is proposed to facilitate efficient predictive coding for spatial redundancy of individual frames. Taking advantage of the relationship between consecutive frames, the SP predictive coding combines with inter-frame differential predictive coding to minimize the intensity of each pixel. Finally, after designing a specific mapping function, a finite length Golomb-Rice encoding is chosen to achieve lossless data compression (Rice & Plaunt 2003). Among them, all steps except for the extraction are reversible; during decompression, corresponding invertible operations are performed to restore the data. For the step involving filling, values outside the solar disk can be assigned to zero, thereby realizing lossless compression of the solar disk information.

3.1. Extraction and Dimensionality Reduction of Solar Disk Information

As we know, the position of the solar disk in the image is fixed, and the largest inscribed circle in the full-disk solar image can be used as a binary template to extract information of the Sun and its outer edge, which can significantly reduce the data volume.

Let the radius of the largest inscribed circle in image be R_{Cut} , then the extracted disk size S_{Cut} is πR_{Cut}^2 , the side length of the original square image is $2R_{\text{Cut}}$, and the size of the square image S_{Image} is $(2R_{\text{Cut}})^2$. The compression ratio $\text{Ratio}_{\text{Cut}}$ is calculated as follows:

$$\text{Ratio}_{\text{Cut}} = \frac{1}{\text{Rate}_{\text{Cut}}} = \frac{S_{\text{Image}}}{S_{\text{Cut}}} = \frac{(2R_{\text{Cut}})^2}{\pi R_{\text{Cut}}^2} = \frac{4}{\pi} \quad (2)$$

The compression ratio $\text{Ratio}_{\text{Cut}}$ is 1.27324, and compression rate Rate_{Cut} is 0.78540. It means the overall data size will be greatly decreased by 22.46%, while not affecting the spatiotemporal continuity of the compressed data.

The remained information comprises two parts: information within the solar disk, and background on the outer edge of the solar disk in the form of a narrow ring (marked in blue in Figure 4). The information on the ring ensures the effectiveness of the proposed method even if the solar disk center deviates from the image center.

Before subsequent predictive coding, the dimensionality reduction process is given as follows: first, choose the horizontal as the reference direction. Then, the circle is dimensionally reduced row by row along the reference direction. It can be expressed as Algorithm 1.

Algorithm 1. Dimensionality Reduction (D-R) or Dimensionality Increase (D-I)

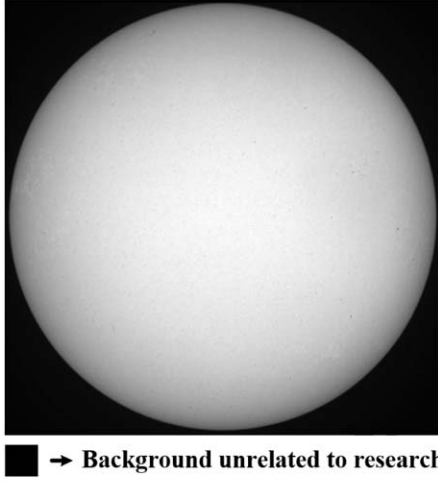
Require: Input $T(u, v)$ and $P(u, v)$. // for D-I, $P(x)$ is required instead of $P(u, v)$.

- 1: Initialize $x \leftarrow 1$.
- 2: **for** $u = 1$ to W **do**
- 3: **for** $v = 1$ to H **do**
- 4: **if** $T(u, v) = 1$ **then**
- 5: $P(x) = P(u, v)$. // for D-I, $P(u, v) = P(x)$.
- 6: $x = x + 1$.
- 7: **end if**
- 8: **end for**
- 9: **end for**

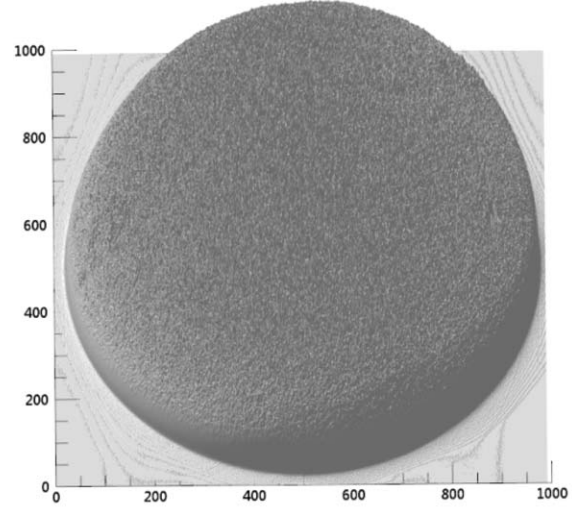
return $P(x)$. // for D-I, return $P(u, v)$.

We defined the $\stackrel{T(u,v)}{\Leftarrow}$ and $\stackrel{T(u,v)}{\Leftarrow}$ as the dimensionality reduction and increase operations:

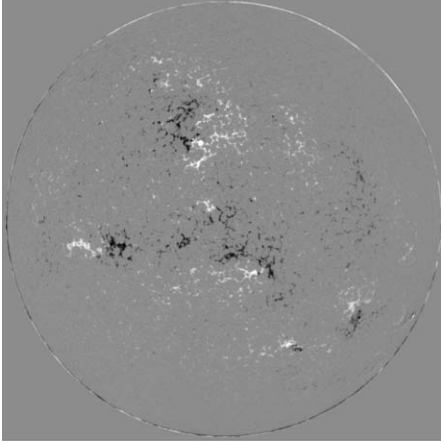
$$\begin{aligned} P^{\ominus}(x) &\stackrel{T(u,v)}{\Leftarrow} P(u, v) \\ P(u, v) &\stackrel{T(u,v)}{\Leftarrow} P^{\ominus}(x) \end{aligned} \quad (3)$$



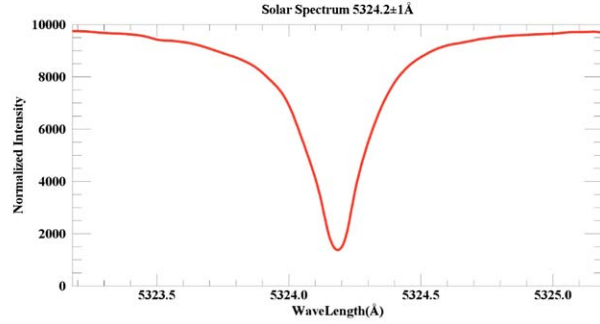
(a) The full-disk solar image



(b) The 3D shape of the intensity of the solar disk



(c) The weak intensity in most area



(d) Spectral intensity at different wavelength

Figure 2. Partial prior knowledge in MHI.

where $P^\ominus(x)$ indicates the data after dimensionality reduction, x denotes its indices, $T(u, v)$ is the template defined by

$$T(u, v) = \begin{cases} 0 & \left[\sqrt{\left(u - \frac{W}{2}\right)^2 + \left(v - \frac{H}{2}\right)^2} \right] \geq R_{\text{Cut}} \\ 1 & \left[\sqrt{\left(u - \frac{W}{2}\right)^2 + \left(v - \frac{H}{2}\right)^2} \right] < R_{\text{Cut}} \end{cases} \quad (4)$$

where $[A]$ represents the calculation of the downward rounding for float A .

3.2. Predictive Coding Design

We design a predictive encoding algorithm for dimensionality reduced data based on the measurement mode and data characteristics of MHI, as shown in Figure 1: For the first frame $P^\ominus(\lambda_c + \Delta\lambda_0, I + Q)$, we utilize the SP predictive coding. For all frames of $P^\ominus(\lambda_c + \Delta\lambda, I - S)$, we employ the inter-frame difference predictive coding between $P^\ominus(\lambda_c + \Delta\lambda, I + S)$ and $P^\ominus(\lambda_c + \Delta\lambda, I - S)$ (R-IFD). For $P^\ominus(\lambda_c + \Delta\lambda, I + U)$ and $P^\ominus(\lambda_c + \Delta\lambda, I + V)$ at the same wavelength position, we select the inter-frame differential predictive coding at the same wavelength (SUVL-IFD). For $P^\ominus(\lambda_c + \Delta\lambda, I + Q)$ except the first frame $P^\ominus(\lambda_c + \Delta\lambda_0, I + Q)$, we apply the inter-frame differential predictive coding in spectral scanning (SQL-IFD).

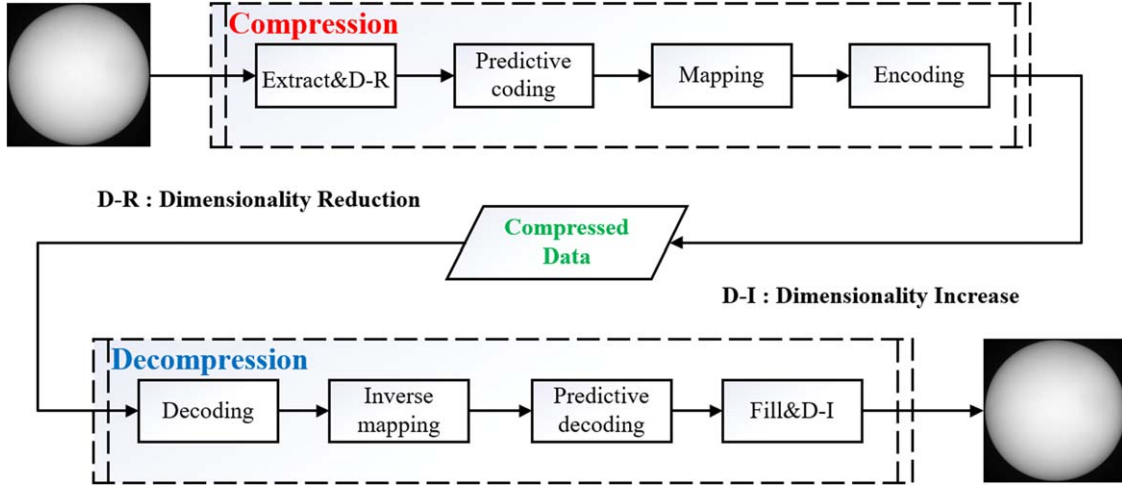


Figure 3. Compression and decompression process.

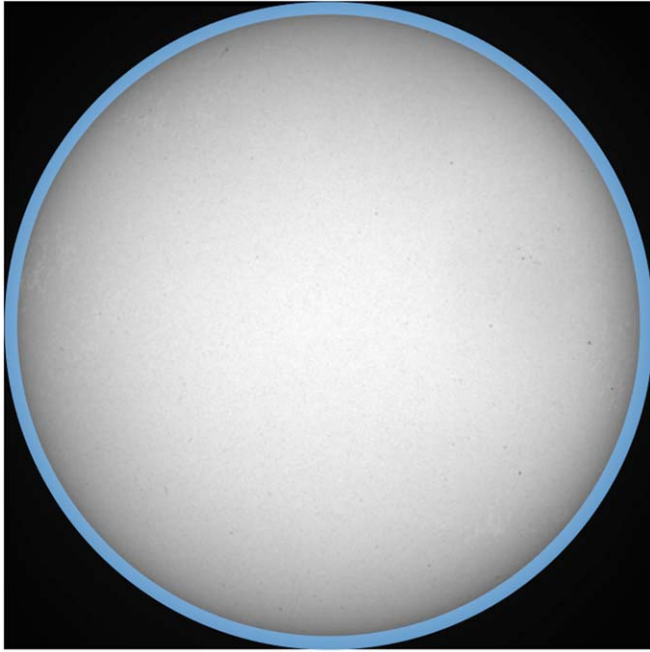


Figure 4. Schematic diagram of the remained information.

3.2.1. The SP Predictive Coding

The solar data contains large-scale spatial redundancy introduced by the limb darkening effect, instrument non-uniformity, etc. on the solar disk, and small-scale spatial redundancy introduced by fine structures such as granules, dark lanes, detector's local noise, cosmic ray noise, etc. Because of the high time cost of removing small-scale spatial redundancy, we only consider large-scale spatial redundancy.

The classic forward pixel predictive coding can remove the spatial redundancy in the reference direction (horizontal) after dimensional reduction of the original image, but there is still spatial redundancy that can be removed in the other direction. In order to remove this redundancy, the forward pixel predictive residual is filled back to the corresponding position of the binary template as a primary predictive residual matrix, which is subsequently transposed. Then, the transposed matrix is reduced in dimensionality again, keeping the reference direction unchanged. After the second dimensionality reduction, the forward pixel predictive coding is performed again, and the final predictive residual is the SP predictive residual. The entire process is referred to as the SP predictive coding, and its calculation process is as follows:

$$\text{Step 1. } P_{PP}^{\odot}(x) = P^{\odot}(x) - P^{\odot}(x - 1)$$

$$\text{Step 2. } P_{PP}(u, v) \stackrel{T(u,v)}{\Leftarrow} P_{PP}^{\odot}(x)$$

$$\text{Step 3. } P_T(u, v) = (P_{PP}(u, v))^{\text{Transpose}}$$

$$\text{Step 4. } P_T^{\odot}(x) \stackrel{T(u,v)}{\Leftarrow} P_T(u, v)$$

$$\text{Step 5. } P_{SP}^{\odot}(x) = P_T^{\odot}(x) - P_T^{\odot}(x - 1)$$

In the process, $P_{PP}^{\odot}(x)$ represents the primary pre-pixel predictive residual, $P_T^{\odot}(x)$ serves as the second dimensionality reduction information, $P_{SP}^{\odot}(x)$ denotes the secondary pre-pixel predictive residual, $P_{PP}(u, v)$ stands for the filled data, and $P_T(u, v)$ serves as the transposed data.

The SP predictive coding can be used both in the predictive coding for the first frame data and in the preprocessing step for subsequent operations. For convenience, the operation of SP is represented by \odot , and the result after the operation is represented by $P^{\odot}(x)$. A dimensionality reduction image after

predicted by SP becomes:

$$P^\ominus(x) = \odot (P^\ominus(x)) \quad (5)$$

Due to $P^\ominus(\lambda_c + \Delta\lambda_0, I + Q)$ being the first frame with no reference, its predictive coding can only be performed based on its own. Therefore, $P^\ominus(\lambda_c + \Delta\lambda_0, I + Q)$ can be expressed as:

$$R(\lambda_c + \Delta\lambda_0, I + Q) = \odot (P^\ominus(\lambda_c + \Delta\lambda_0, I + Q)), \quad (6)$$

where $R(\lambda_c + \Delta\lambda_0, I + Q)$ is the residual after SP predictive coding.

3.2.2. Inter Frame Differential Predictive Coding

All polarization parameters in most areas of the solar disk are weak. It can remove the spatiotemporal redundancy of $P^\ominus(I - S)$ by using the inter-frame difference predictive coding between $P^\ominus(I + S)$ and $P^\ominus(I - S)$ (R-IFD):

$$R(I - S) = P^\ominus(I + S) - P^\ominus(I - S) \quad (S = Q, U, V). \quad (7)$$

where $R(I - S)$ is the residual after R-IFD predictive coding.

In addition, the sum of $P(I + S)$ and $P(I - S)$ at the same wavelength is approximately equal. By applying the SP predictive coding and Equation (8), the spatiotemporal redundancy of $P^\ominus(I + U)$ and $P^\ominus(I + V)$ can be eliminated. This is the inter-frame differential predictive coding at the same wavelength (SUVL-IFD).

$$\begin{aligned} R(I + U) &= \lfloor (P^\ominus(I + Q) + P^\ominus(I - Q))/2 \rfloor \\ &\quad - P^\ominus(I + U) \\ R(I + V) &= \lfloor (P^\ominus(I + U) + P^\ominus(I - U))/2 \rfloor \\ &\quad - P^\ominus(I + V) \end{aligned} \quad (8)$$

where $R(I + U)$, $R(I + V)$ are the results of eliminating spatiotemporal redundancy.

In fact that the $P^\ominus(\lambda_c + \Delta\lambda, I + Q)$ exhibits large-scale changes at different wavelength positions, we employ the SP predictive coding to smooth out the large-scale changes in image intensity and eliminate the influence of image jitter. Then, we normalize the intensity of the previous frame to the current one, perform SP predictive coding on both frames, and differentiate their results after predictive coding. This process is defined as the inter-frame differencing predictive coding in spectral scanning (SQL-IFD). How we eliminate the spatiotemporal redundancy of adjacent images in spectral scanning is as follows:

$$\begin{aligned} R(\lambda_c + \Delta\lambda_j, I + Q) &= \odot \left[\frac{\overline{P^\ominus(\lambda_c + \Delta\lambda_j, I + Q)}}{\overline{P^\ominus(\lambda_c + \Delta\lambda_{j-1}, I + Q)}} \right. \\ &\quad \cdot P^\ominus(\lambda_c + \Delta\lambda_{j-1}, I + Q) \left. \right] - P^\ominus(\lambda_c + \Delta\lambda_j, I + Q) \end{aligned} \quad (9)$$

where \overline{B} represents the calculation of the mean of matrix B , and j is the index of different wavelength.

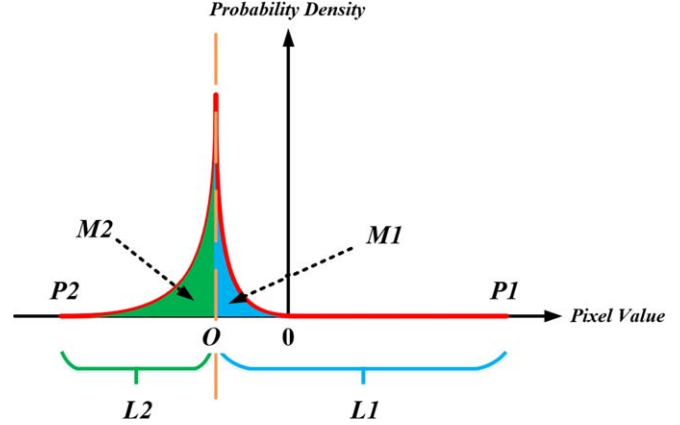


Figure 5. The general normal form of probability density distribution for predictive residual data.

3.3. Mapping and Encoding Design

The mapped data in the article approximately follow a univariate geometric distribution. Considering the usability requirements for the encoding algorithm in space projects (i.e., no patent issues, low complexity, and simple hardware implementation), the Golomb-Rice algorithm is selected for encoding.

To employ the Golomb-Rice coding, an ideal source should follow a discrete geometric distribution (Merhav et al. 1998). Typically, the probability density distribution of the predicted residuals exhibits slight asymmetry, as shown in Figure 5. The vertical line through the peak point serves as the approximate symmetry axis: the long side extends to $P1$, the length from the peak point to $P1$ is $L1$, and the blue area is $M1$; the short side extends to $P2$, the length from the peak point to $P2$ is $L2$, and the green area is $M2$.

The distribution in Figure 5 deviates from the ideal geometric distribution, so it is necessary to map the existing predictive residual data set from a general normal form to an approximate discrete geometric distribution.

The final mapping equation is as follows:

$$R' = \begin{cases} -2(R - O) - 1 & (R - O) < 0 & M2 > M1 & |R - O| \leq L2 \\ 2(R - O) & (R - O) \geq 0 & M2 > M1 & |R - O| \leq L2 \\ -2(R - O) & (R - O) \leq 0 & M2 \leq M1 & |R - O| \leq L2 \\ 2(R - O) - 1 & (R - O) > 0 & M2 \leq M1 & |R - O| \leq L2 \\ |R - O| + |L2| & \forall M1, \forall M2 & |R - O| > L2 \end{cases} \quad (10)$$

where R is the predictive residual, R' is the mapped result, $M1/M2$ represents the cumulative probability density for residual data, which is calculated by integrating the probability density from the peak point to $P1/P2$.

The entire lossless compression method can be summarized as following steps:

1. Reduce all data using Equation (3).

2. According to Figure 1, different data are predictively coded using corresponding methods. For details, refer to Equations (6)–(9).
3. Map the predictive residual according to Equation (10), and apply finite-length Golomb-Rice coding to the result.

4. Experiment

4.1. Data Selection and Experimental Design

The Solar Full-disk Multi-layer Magnetograph (SFMM) at Ganyu station has a similar measurement mode and data characteristics as MHI. Therefore, the SFMM's data are used in our experiments to test the compression method. A group of spectral scanning data by SFMM includes Stokes $I \pm Q$, $I \pm U$, $I \pm V$ measurements at 6 wavelength positions, resulting in a total of 18 files, the bit depth of them is $32 \text{ bit pixel}^{-1}$.

The characteristics of an image, its noise level, and the compression method employed together determine the magnitude of data compression ratio. Under routine scientific observation modes, MHI's data is not affected by cloud cover; hence, when the signal-to-noise ratio of the detector remains constant, the primary variable among image characteristics affecting the data compression ratio is the variation in the number of sunspots. To guarantee the robustness of the method, we select three groups of SFMM's data with different sunspot characteristics as the test set, including no sunspots, small sunspots, and sunspot groups. There are 54 files in total, each containing two images, making it a total of 108 images. The compression ratio of these files can serve as evidence for the universality of this method.

The experiment is divided into three parts:

1. Compress all the data in the test set and analyze whether the compression ratio meets the MHI's requirements.
2. Compress one group of the data sets using multiple compression algorithms, and analyze whether the compression ratio in this article is superior to the general algorithms.
3. Analyze whether the time consumption of our compression method satisfies the MHI's real-time requirement.

4.2. Result Analysis

4.2.1. Analysis of Compression Ratio

The compression ratios for the three sets of spectral scanning data by our lossless method are shown in Figure 6. Regardless of the varying number of sunspots, a consistent trend is observed in the overall distribution of compression ratios, with an average ratio of approximately 3.67 and a standard deviation less than 0.11 (shown in Table 1). This indicates that not only does the method significantly surpass the SPO requirement of 3.5, but it also exhibits small fluctuations, suggesting minimal

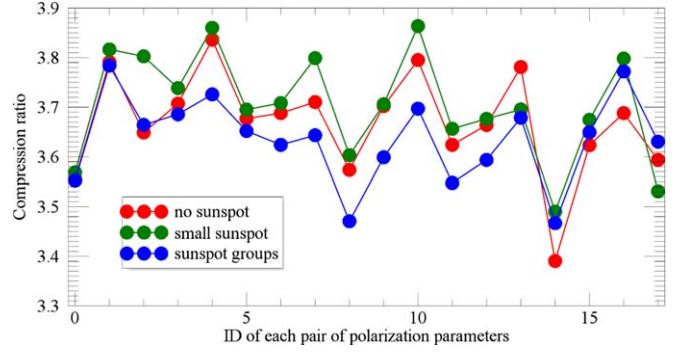


Figure 6. Compression ratio of the test set.

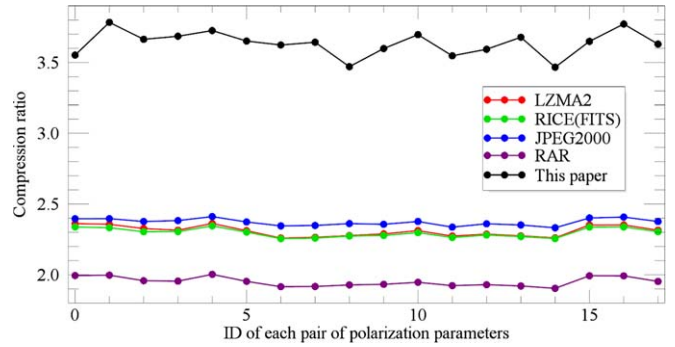


Figure 7. Comparison with common lossless compression algorithms.

Table 1
Statistics of Compression Ratio of Spectral Scanning Data

	No Sunspot	Small Sunspot	Sunspot Groups
Mean value	3.669 406 4	3.704 568 6	3.635 353 7
Standard deviation	0.104 567 26	0.108 322 20	0.088 045 018

influence from the number of sunspots. Thus, the method demonstrates a strong universality.

4.2.2. Comparison with Other Lossless Compression Algorithms

To demonstrate the compression performance of the method proposed in this article, we compared JPEG2000, which is widely used in spatial compression, the partition RICE compression algorithm taken by the Flexible Image Transport System (FITS), and the mature compression algorithms lzma2 and RAR. The result of compressing the raw data (including solar disk and all background information) is shown in Figure 7:

From Figure 7, it is evident that among the alternative methods compared to ours, RAR has the lowest compression ratio, followed by RICE in FITS, LZMA2 compression, and JPEG2000, with their best compression ratio approaching

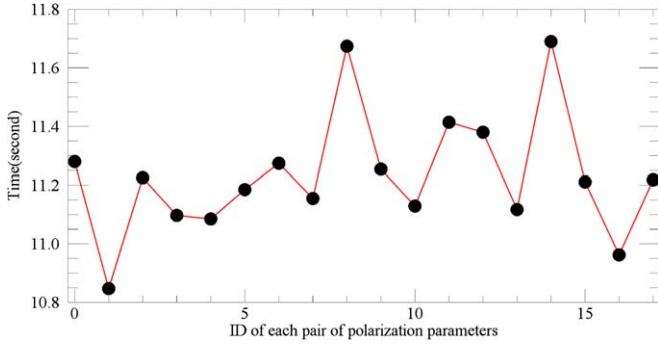


Figure 8. The compression time for each pair of polarization parameters.

approximately 2.5. In contrast, our proposed method achieves the highest compression ratio, surpassing 3.5. Even if these contrastive methods were to concentrate exclusively on compressing the solar disk, according to Equation (2), their optimal compression ratio would theoretically approach 3.1831, which is still less than 3.5. It demonstrates that our method is particularly more suitable for lossless compression of the solar full-disk Stokes images in the spectral scanning mode. Note that our method has the same trend with others except for slightly big fluctuation, because we retained only the solar disk data during dimensionality reduction.

4.2.3. Time Consumption of Compression Method

Nowadays Loogson processors have been successfully applied in space projects of China for onboard processing, and the performance of the latest Loogson 3A6000 processor (with a clock frequency of 2.0–2.5GHz) is similar to that of the 10th generation Intel Core processor. So in this paper we use a single core of the 10th generation Core i7 10750H processor instead of the Loogson 3A6000 processor for the time consumption testing, with Turbo Boost disabled and the peak frequency limited to no more than 2.28 GHz. The test data set is from sunspot groups.

Figure 8 shows that the compression time for a pair of $P(I \pm S)$ is less than 12 s. A single observation cycle of the MHI payload will last 15 minutes, and there will be a total of 18 pairs of $P(I \pm S)$. Therefore, using the compression method proposed in this paper, the compression time is less than 216 s, which is much shorter than the observation cycle of 15 minutes, meeting the time requirements for MHI.

5. Discussion and Conclusion

In order to address the urgent need for efficient data lossless compression in MHI, this article proposes a targeted lossless compression method based on the measurement mode and data characteristics. The main process and test results of our method are as follows: The background information outside the solar disk is extracted and dimensionality reduced by utilizing the largest inscribed circle within the image, which results in an

overall data volume reduction of no less than 22.46%; subsequently, predictive coding is performed to eliminate all correlations (space, spectrum, and polarization) in solar data cube: a single frame is predicted using SP predictive coding, while the remaining frames are predicted using inter-frame differential predictive encoding combined with SP predictive coding. After mapping and finite length Golomb-Rice encoding, the final average compression ratio is up to 3.67. It is significantly improved compared with general lossless methods such as LZMA2, RAR, RICE in FITS, and JPEG2000. At the same time, the average processing time of our method for a set of data is less than one observation period, meeting real-time requirements. Due to the limited capability of the acquisition device and processor, our method only involves data binarization, matrix multiplication, transposition, differentiation, array traversal, etc. The subsequent encoding is a mature scheme that has been in general use for many years. Therefore, the proposed method has strong feasibility, and can be quickly ported to the payload for testing its practical time consumption and effect.

Acknowledgments

We appreciate the staff of Ganyu station providing these excellent data, and thank the staff at the Huairou Solar Observing Station for their help. This study is supported by the National Key R&D Program of China (grant No.2022YFF0503800); by the National Natural Science Foundation of China (NSFC) (grant No.11427901); by the Strategic Priority Research Program of the Chinese Academy of Sciences (CAS-SPP) (grant No. XDA15320102); by the Youth Innovation Promotion Association (CAS No.2022057)

ORCID iDs

Kai-Fan Ji  <https://orcid.org/0000-0001-8950-3875>

References

- Ai, G.-X., & Hu, Y.-F. 1986, *AcASn*, 86, 91
- CCSDS 2017, Recommendation for Space Data System Standards (Blue Book) (Washington, DC: CCSDS Secretariat)
- CCSDS 2020, Recommendation for Space Data System Standards (Blue Book) (Washington, DC: CCSDS Secretariat)
- Deng, Y.-Y., Zhang, H.-Y., Yang, J.-F., et al. 2019, *RAA*, 19, 157
- Deng, Y.-Y., Zhou, G.-P., Dai, S.-W., et al. 2023, *ChSBu*, 68, 298
- Fischer, C. E., Müller, D., & Moortel, I. D. 2016, *SoPh*, 292, 16
- Golomb, S. 1966, *ITIT*, 12, 399
- ITU-T 1998, Information technology-Lossless and near-lossless compression of continuous-tone still images: Baseline, ITU-T Recommendation T., 87, 1
- ITU-T 2002, Information technology – JPEG 2000 image coding system: Core coding system, ITU-T Recommendation T., 800, 1
- Liu, Y. 2018, Lossless Compression Algorithm and Key Technologies for Full-Surface Solar Magnetic Field Images, Master's Professional Degree Thesis, Beijing Jiaotong University
- Merhav, N., Seroussi, G., & Weinberger, M. J. 1998, *ITIT*, 46, 229
- Müller, D., Cyr, O. C. S., Zouganelis, I., et al. 2020, *A&A*, 642, A1
- Nandy, D., Banerjee, D., Bhowmik, P., et al. 2023, *BAAS*, 55, 287
- Rice, R., & Plaunt, J. 2003, *IEEE Trans. Commun. Technol.*, 19, 889
- Wenzel, K. P., Marsden, R. G., Page, D. E., et al. 1992, *A&AS*, 92, 207
- Zhang, H.-Q., Wang, D.-G., Deng, Y.-Y., et al. 2007, *ChJAA*, 7, 281